EAST Search History

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	27	cluster and margin and unsupervised	USPAT	OR	OFF	2006/04/30 23:34
L2	98	cluster and margin and unsupervised	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2006/04/30 23:34

Results for "((cluster and margin and unsupervised)<in>metadata)" e-mail

Your search matched 1 of 1344017 documents.

A maximum of 100 results are displayed, 25 to a page, sorted by Relevance in Descending order.

» Search Options

View Session History

New Search

Modify Search

((cluster and margin and unsupervised)<in>metadata) Search

Check to search only within this results set

» Key

Display Format: Citation Citation & Abstract

IEEE JNL	IEEE Journal or Magazine
IEE JNL	IEE Journal or Magazine
IEEE CNF	IEEE Conference Proceeding
IEE CNF	IEE Conference Proceeding
IEEE STD	IEEE Standard

view selected items Select All Deselect All

1. **A new kernel clustering algorithm**
   Borer, S.; Gerstner, W.;
   Neural Information Processing, 2002. ICONIP '02. Proceedings of the 9th International Conference
   Volume 5, 18-22 Nov. 2002 Page(s):2527 - 2531 vol.5
   Digital Object Identifier 10.1109/ICONIP.2002.1201950
   AbstractPlus | Full Text: PDF(492 KB) IEEE CNF
   Rights and Permissions

Terms used **cluster** and **margin** and **unsupervised** Found **3,393** of **175,083**

Sort results by: relevance
Display results: expanded form
Save results to a Binder
Search Tips
Open results in a new window

Try an Advanced Search
Try this search in The ACM Guide

Results 1 - 20 of 200 Result page: **1** 2 3 4 5 6 7 8 9 10 next
Best 200 shown Relevance scale

1 Associative Clustering for Exploring Dependencies between Functional Genomics Data Sets
Samuel Kaski, Janne Nikkila, Janne Sinkkonen, Leo Lahti, Juha E. A. Knuuttila, Christophe Roos
July 2005 **IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)**, Volume 2 Issue 3
Publisher: IEEE Computer Society Press
Full text available: pdf(896.56 KB) Additional Information: full citation, abstract, references, index terms

High-throughput genomic measurements, interpreted as cooccurring data samples from multiple sources, open up a fresh problem for machine learning: What is in common in the different data sets, that is, what kind of statistical dependencies are there between the paired samples from the different sets? We introduce a clustering algorithm for exploring the dependencies. Samples within each data set are grouped such that the dependencies between groups of different sets capture as much of pairwise d ...

Keywords: Index Terms- Biology and genetics, clustering, contingency table analysis, machine learning, multivariate statistics.

2 Variational learning of clusters of undercomplete nonsymmetric independent components
Kwokleung Chan, Te-Won Lee, Terrence J. Sejnowski
March 2003 **The Journal of Machine Learning Research**, Volume 3
Publisher: MIT Press
Full text available: pdf(345.39 KB) Additional Information: full citation, abstract, references, citings, index terms

We apply a variational method to automatically determine the number of mixtures of independent components in high-dimensional datasets, in which the sources may be nonsymmetrically distributed. The data are modeled by clusters where each cluster is described as a linear mixture of independent factors. The variational Bayesian method yields an accurate density model for the observed data without overfitting problems. This allows the dimensionality of the data to be identified for each cluster. Th ...

Keywords: Bayesian learning, ICA, density estimations, mixture models

3 Feature Selection for Unsupervised Learning
Jennifer G. Dy, Carla E. Brodley
December 2004 **The Journal of Machine Learning Research**, Volume 5
Publisher: MIT Press

Full text available: pdf(725.21 KB) Additional Information: full citation, abstract

In this paper, we identify two issues involved in developing an automated feature subset selection algorithm for unlabeled data: the need for finding the number of clusters in conjunction with feature selection, and the need for normalizing the bias of feature selection criteria with respect to dimension. We explore the feature selection problem and these issues through FSSEM (Feature Subset Selection using Expectation-Maximization (EM) clustering) and through two different performance criteria ...

4 Text Categorization: Unsupervised document classification using sequential information maximization

Noam Slonim, Nir Friedman, Naftali Tishby

August 2002 **Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval**

Publisher: ACM Press

Full text available: pdf(236.71 KB) Additional Information: full citation, abstract, references, citings, index terms

We present a novel sequential clustering algorithm which is motivated by the *Information Bottleneck (IB)* method. In contrast to the agglomerative *IB* algorithm, the new sequential (*sIB*) approach is guaranteed to converge to a local maximum of the information with time and space complexity typically linear in the data size. information, as required by the original IB principle. Moreover, the time and space complexity are significantly improved. We apply this algorithm to unsup ...

5 Machine learning in automated text categorization

Fabrizio Sebastiani

March 2002 **ACM Computing Surveys (CSUR)**, Volume 34 Issue 1

Publisher: ACM Press

Full text available: pdf(524.41 KB) Additional Information: full citation, abstract, references, citings, index terms

The automated categorization (or classification) of texts into predefined categories has witnessed a booming interest in the last 10 years, due to the increased availability of documents in digital form and the ensuing need to organize them. In the research community the dominant approach to this problem is based on machine learning techniques: a general inductive process automatically builds a classifier by learning, from a set of preclassified documents, the characteristics of the categories. ...

Keywords: Machine learning, text categorization, text classification

6 Compression and summarization: Supervised ranking in open-domain text summarization

Tadashi Nomoto, Yuji Matsumoto

July 2001 **Proceedings of the 40th Annual Meeting on Association for Computational Linguistics ACL '02**

Publisher: Association for Computational Linguistics

Full text available: pdf(142.01 KB) Additional Information: full citation, abstract, references

The paper proposes and empirically motivates an integration of supervised learning with unsupervised learning to deal with human biases in summarization. In particular, we explore the use of probabilistic decision tree within the clustering framework to account for the variation as well as regularity in human created summaries. The corpus of human created extracts is created from a newspaper corpus and used as a test set. We build probabilistic decision trees of different flavors and integrate e ...

7 Clustering: Restrictive clustering and metaclustering for self-organizing document collections

Stefan Siersdorfer, Sergej Sizov

July 2004 **Proceedings of the 27th annual international ACM SIGIR conference on**

Research and development in information retrieval SIGIR '04

Publisher: ACM Press

Full text available: pdf(171.71 KB) Additional Information: full citation, abstract, references, index terms

This paper addresses the problem of automatically structuring heterogenous document collections by using clustering methods. In contrast to traditional clustering, we study restrictive methods and ensemble-based meta methods that may decide to leave out some documents rather than assigning them to inappropriate clusters with low confidence. These techniques result in higher cluster purity, better overall accuracy, and make unsupervised self-organization more robust. Our comprehensive experimenta ...

Keywords: meta clustering, restrictive clustering

8 Tissue classification with gene expression profiles

Amir Ben-Dor, Laurakay Bruhn, Nir Friedman, Iftach Nachman, Michèl Schummer, Zohar Yakhini

April 2000 **Proceedings of the fourth annual international conference on Computational molecular biology**

Publisher: ACM Press

Full text available: pdf(1.11 MB) Additional Information: full citation, abstract, references, citings

Constantly improving gene expression profiling technologies are expected to provide understanding and insight into cancer related cellular processes. Gene expression data is also expected to significantly and in the development of efficient cancer diagnosis and classification platforms. In this work we examine two sets of gene expression data measured across sets of tumor and normal clinical samples One set consists of 2,000 genes, measured in 62 epithelial colon samples [1]. The second consi ...

9 Boosting margin based distance functions for clustering

Tomer Hertz, Aharon Bar-Hillel, Daphna Weinshall

July 2004 **Proceedings of the twenty-first international conference on Machine learning ICML '04**

Publisher: ACM Press

Full text available: pdf(181.49 KB) Additional Information: full citation, abstract, references

The performance of graph based clustering methods critically depends on the quality of the distance function used to compute similarities between pairs of neighboring nodes. In this paper we learn distance functions by training binary classifiers with margins. The classifiers are defined over the product space of pairs of points and are trained to distinguish whether two points come from the same class or not. The signed margin is used as the distance value. Our main contribution is a distance l ...

10 A new approach to unsupervised text summarization

Tadashi Nomoto, Yuji Matsumoto

September 2001 **Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval**

Publisher: ACM Press

Full text available: pdf(215.56 KB) Additional Information: full citation, abstract, references, citings, index terms

The paper presents a novel approach to unsupervised text summarization. The novelty lies in exploiting the diversity of concepts in text for summarization, which has not received much attention in the summarization literature. A diversity-based approach here is a principled generalization of Maximal Marginal Relevance criterion by Carbonell and Goldstein \cite{carbonell-goldstein98}.We propose, in addition, aninformation-centricapproach to evaluation, where the q ...

Keywords: text summarization

11 Think globally, fit locally: unsupervised learning of low dimensional manifolds

Lawrence K. Saul, Sam T. Roweis

December 2003 **The Journal of Machine Learning Research**, Volume 4

Publisher: MIT Press

Full text available: pdf(2.91 MB) Additional Information: full citation, abstract, references, citings, index terms

The problem of dimensionality reduction arises in many fields of information processing, including machine learning, data compression, scientific visualization, pattern recognition, and neural computation. Here we describe locally linear embedding (LLE), an unsupervised learning algorithm that computes low dimensional, neighborhood preserving embeddings of high dimensional data. The data, assumed to be sampled from an underlying manifold, are mapped into a single global coordinate system of lowe ...

12 Research track: New unsupervised clustering algorithm for large datasets

William Peter, John Chiochetti, Clare Giardina

August 2003 **Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining**

Publisher: ACM Press

Full text available: pdf(12.53 MB) Additional Information: full citation, abstract, references, index terms

A fast and accurate unsupervised clustering algorithm has been developed for clustering very large datasets. Though designed for very large volumes of geospatial data, the algorithm is general enough to be used in a wide variety of domain applications. The number of computations the algorithm requires is ~ $O(N)$, and thus faster than hierarchical algorithms. Unlike the popular K-means heuristic, this algorithm does not require a series of iterations to converge to a solution. In add ...

Keywords: clustering, data streaming, geospatial data, large datasets

13 Solving regression problems with rule-based ensemble classifiers

Nitin Indurkhya, Sholom M. Weiss

August 2001 **Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining**

Publisher: ACM Press

Full text available: pdf(556.71 KB) Additional Information: full citation, abstract, references, citings, index terms

We describe a lightweight learning method that induces an ensemble of decision-rule solutions for regression problems. Instead of direct prediction of a continuous output variable, the method discretizes the variable by k-means clustering and solves the resultant classification problem. Predictions on new examples are made by averaging the mean values of classes with votes that are close in number to the most likely class. We provide experimental evidence that this indirect approach can often yi ...

14 Unsupervised induction of stochastic context-free grammars using distributional clustering

Alexander Clark

July 2001 **Proceedings of the 2001 workshop on Computational Natural Language Learning - Volume 7 ConLL '01**

Publisher: Association for Computational Linguistics

Full text available: pdf(99.58 KB) Additional Information: full citation, abstract, references

An algorithm is presented for learning a phrase-structure grammar from tagged text. It clusters sequences of tags together based on local distributional information, and selects clusters that satisfy a novel mutual information criterion. This criterion is shown to be related to the entropy of a random variable associated with the tree structures, and it is demonstrated that it selects linguistically plausible constituents. This is incorporated in a Minimum Description Length algorithm. The evalu ...

15 Articles on microarray data mining: Machine learning in low-level microarray analysis

Benjamin I. P. Rubinstein, Jon McAuliffe, Simon Cawley, Marimuthu Palaniswami, Kotagiri Ramamohanarao, Terence P. Speed

December 2003 **ACM SIGKDD Explorations Newsletter**, Volume 5 Issue 2

Publisher: ACM Press

Full text available: pdf(382.35 KB) Additional Information: full citation, abstract, references

Machine learning and data mining have found a multitude of successful applications in microarray analysis, with gene clustering and classification of tissue samples being widely cited examples. Low-level microarray analysis -- often associated with the pre-processing stage within the microarray life-cycle -- has increasingly become an area of active research, traditionally involving techniques from classical statistics. This paper explores opportunities for the application of machine learning an ...

Keywords: gene expression estimation, genotyping, incremental learning, learning from heterogeneous data, low-level microarray analysis, re-sequencing, semi-supervised learning, transcript discovery, transductive learning

16 Special issue on special feature: Distributional word clusters vs. words for text categorization

Ron Bekkerman, Ran El-Yaniv, Naftali Tishby, Yoad Winter

March 2003 **The Journal of Machine Learning Research**, Volume 3

Publisher: MIT Press

Full text available: pdf(176.53 KB) Additional Information: full citation, abstract, index terms

We study an approach to text categorization that combines distributional clustering of words and a Support Vector Machine (SVM) classifier. This word-cluster representation is computed using the recently introduced *Information Bottleneck* method, which generates a compact and efficient representation of documents. When combined with the classification power of the SVM, this method yields high performance in text categorization. This novel combination of SVM with word-cluster representation ...

17 Semisupervised Learning for Molecular Profiling

Cesare Furlanello, Maria Serafini, Stefano Merler, Giuseppe Jurman

April 2005 **IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)**, Volume 2 Issue 2

Publisher: IEEE Computer Society Press

Full text available: pdf(1.09 MB) Additional Information: full citation, abstract, references, index terms

Class prediction and feature selection are two learning tasks that are strictly paired in the search of molecular profiles from microarray data. Researchers have become aware how easy it is to incur a selection bias effect, and complex validation setups are required to avoid overly optimistic estimates of the predictive accuracy of the models and incorrect gene selections. This paper describes a semisupervised pattern discovery approach that uses the by-products of complete validation studies on ...

Keywords: Machine learning, data mining, classifier design and evaluation, feature evaluation and selection, pattern analysis, clustering, similarity measures, biology and genetics, bioinformatics databases.

18 A cross-comparison of two clustering methods

Olivier Ferret, Brigitte Grau, Michèle Jardino

July 2001 **Proceedings of the workshop on Evaluation for Language and Dialogue Systems - Volume 9**

Publisher: Association for Computational Linguistics

Full text available: pdf(85.99 KB) Additional Information: full citation, abstract, references

Many Natural Language Processing applications require semantic knowledge about topics in order to be possible or to be efficient. So we developed a system, SEGAPSITH, that

acquires it automatically from text segments by using an unsupervised and incremental clustering method. In such an approach, an important problem consists of the validation of the learned classes. To do that, we applied another clustering method, that only needs to know the number of classes to build, on the same subset of te ...

19 Special issue on kernel methods: Support vector clustering

Asa Ben-Hur, David Horn, Hava T. Siegelmann, Vladimir Vapnik
March 2002 **The Journal of Machine Learning Research**, Volume 2
Publisher: MIT Press
Full text available: pdf(343.44 KB) Additional Information: full citation, abstract, citings

We present a novel clustering method using the approach of support vector machines. Data points are mapped by means of a Gaussian kernel to a high dimensional feature space, where we search for the minimal enclosing sphere. This sphere, when mapped back to data space, can separate into several components, each enclosing a separate cluster of points. We present a simple algorithm for identifying these clusters. The width of the Gaussian kernel controls the scale at which the data is probed while ...

20 Research track: Information-theoretic co-clustering

Inderjit S. Dhillon, Subramanyam Mallela, Dharmendra S. Modha
August 2003 **Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining**
Publisher: ACM Press
Full text available: pdf(1.96 MB) Additional Information: full citation, abstract, references, citings, index terms

Two-dimensional contingency or co-occurrence tables arise frequently in important applications such as text, web-log and market-basket data analysis. A basic problem in contingency table analysis is *co-clustering: simultaneous clustering* of the rows and columns. A novel theoretical formulation views the contingency table as an empirical joint probability distribution of two discrete random variables and poses the co-clustering problem as an optimization problem in *information theory*
Keywords: co-clustering, information theory, mutual information

Terms used **cluster** and **margin** and **unsupervised** Found **3,393** of **175,083**

Sort results by: relevance
Display results: expanded form
Save results to a Binder
Search Tips
Open results in a new window
Try an Advanced Search
Try this search in The ACM Guide

Results 21 - 40 of 200 Result page: previous 1 **2** 3 4 5 6 7 8 9 10 next
Best 200 shown Relevance scale

21 Data clustering: a review

A. K. Jain, M. N. Murty, P. J. Flynn
September 1999 **ACM Computing Surveys (CSUR)**, Volume 31 Issue 3
Publisher: ACM Press
Full text available: pdf(636.24 KB) Additional Information: full citation, abstract, references, citings, index terms, review

Clustering is the unsupervised classification of patterns (observations, data items, or feature vectors) into groups (clusters). The clustering problem has been addressed in many contexts and by researchers in many disciplines; this reflects its broad appeal and usefulness as one of the steps in exploratory data analysis. However, clustering is a difficult problem combinatorially, and differences in assumptions and contexts in different communities has made the transfer of useful generic co ...

Keywords: cluster analysis, clustering applications, exploratory data analysis, incremental clustering, similarity indices, unsupervised learning

22 Paper session IR-3 (information retrieval): web retrieval: Person resolution in person search results: WebHawk

Xiaojun Wan, Jianfeng Gao, Mu Li, Binggong Ding
October 2005 **Proceedings of the 14th ACM international conference on Information and knowledge management CIKM '05**
Publisher: ACM Press
Full text available: pdf(616.03 KB) Additional Information: full citation, abstract, references, index terms

Finding information about people on the Web using a search engine is difficult because there is a many-to-many mapping between person names and specific persons (i.e. referents). This paper describes a person resolution system, called **WebHawk**. Given a list of pages obtained by submitting a person query to a search engine, **WebHawk** facilitates person search in three steps: First of all, a *filter* removes those pages that contain no information about any person. Secondly, a *c* ...

Keywords: *clustering, junk filtering, person resolution, person search*

23 Unsupervised clustering of robot activities: a Bayesian approach

Marco Ramoni, Paola Sebastiani, Paul Cohen
June 2000 **Proceedings of the fourth international conference on Autonomous agents**
Publisher: ACM Press
Full text available: pdf(217.56 KB) Additional Information: full citation, references, citings, index terms

24 Content 2: image clustering: Building and tracking hierarchical geographical & temporal partitions for image collection management on mobile devices

A. Pigeau, M. Gelgon

November 2005 **Proceedings of the 13th annual ACM international conference on Multimedia MULTIMEDIA '05**

Publisher: ACM Press

Full text available: pdf(312.95 KB) Additional Information: full citation, abstract, references, index terms

Usage of mobile devices (phones, digital cameras) raises the need for organizing large personal image collections. In accordance with studies on user needs, we propose a statistical criterion and an associated optimization technique, relying on geo-temporal image metadata, for building and tracking a hierarchical structure on the image collection. In a mixture model framework, particularities of the application and typical data sets are taken into account in the design of the scheme (incremental ...

Keywords: clustering, consumer, database, multimedia

25 Learning to cluster using local neighborhood structure

Rómer Rosales, Kannan Achan, Brendan Frey

July 2004 **Proceedings of the twenty-first international conference on Machine learning ICML '04**

Publisher: ACM Press

Full text available: pdf(318.54 KB) Additional Information: full citation, abstract, references

This paper introduces an approach for clustering/classification which is based on the use of local, high-order structure present in the data. For some problems, this local structure might be more relevant for classification than other measures of point similarity used by popular unsupervised and semi-supervised clustering methods. Under this approach, changes in the class label are associated to changes in the local properties of the data. Using this idea, we also pursue to *learn how to clust* ...

26 Intrusion and privacy: Exploiting unlabeled data in ensemble methods

Kristin P. Bennett, Ayhan Demiriz, Richard Maclin

July 2002 **Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining**

Publisher: ACM Press

Full text available: pdf(719.46 KB) Additional Information: full citation, abstract, references, index terms

An adaptive semi-supervised ensemble method, ASSEMBLE, is proposed that constructs classification ensembles based on both labeled and unlabeled data. ASSEMBLE alternates between assigning "pseudo-classes" to the unlabeled data using the existing ensemble and constructing the next base classifier using both the labeled and pseudolabeled data. Mathematically, this intuitive algorithm corresponds to maximizing the classification margin in hypothesis space as measured on both the labeled and unlabel ...

Keywords: boosting, classification, ensemble learning, semi-supervised learning

27 A hierarchical method for multi-class support vector machines

Volkan Vural, Jennifer G. Dy

July 2004 **Proceedings of the twenty-first international conference on Machine learning ICML '04**

Publisher: ACM Press

Full text available: pdf(171.20 KB) Additional Information: full citation, abstract, references

We introduce a framework, which we call Divide-by-2 (DB2), for extending support vector machines (SVM) to multi-class problems. DB2 offers an alternative to the standard one-against-one and one-against-rest algorithms. For an *N* class problem, DB2 produces an *N*

– 1 node binary decision tree where nodes represent decision boundaries formed by $N - 1$ SVM binary classifiers. This tree structure allows us to present a generalization and a time complexity analysis of DB ...

28 Unsupervised Bayesian visualization of high-dimensional data

Petri Kontkanen, Jussi Lahtinen, Petri Myllymäki, Henry Tirri
August 2000 **Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining**
Publisher: ACM Press
Full text available: pdf(160.91 KB) Additional Information: full citation, references, index terms

29 Posters: Mining multimedia salient concepts for incremental information extraction

João Magalhães, Stefan Rüger
August 2005 **Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval SIGIR '05**
Publisher: ACM Press
Full text available: pdf(161.76 KB) Additional Information: full citation, abstract, references, index terms

We propose a novel algorithm for extracting information by mining the feature space clusters and then assigning salient concepts to them. Bayesian techniques for extracting concepts from multimedia usually suffer either from lack of data or from too complex concepts to be represented by a single statistical model. An incremental information extraction approach, working at different levels of abstraction, would be able to handle concepts of varying complexities. We present the results of our rese ...

Keywords: multimedia clustering, multimedia information extraction

30 Biclustering Models for Structured Microarray Data

Heather L. Turner, Trevor C. Bailey, Wojtek J. Krzanowski, Cheryl A. Hemingway
October 2005 **IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)**, Volume 2 Issue 4
Publisher: IEEE Computer Society Press
Full text available: pdf(1.41 MB) Additional Information: full citation, abstract, index terms

Microarrays have become a standard tool for investigating gene function and more complex microarray experiments are increasingly being conducted. For example, an experiment may involve samples from several groups or may investigate changes in gene expression over time for several subjects, leading to large three-way data sets. In response to this increase in data complexity, we propose some extensions to the plaid model, a biclustering method developed for the analysis of gene expression data. T ...

Keywords: Biclustering, two-way clustering, overlapping clustering, partial supervision, repeated measures, three-way data.

31 Special issue on special feature: A divisive information theoretic feature clustering algorithm for text classification

Inderjit S. Dhillon, Subramanyam Mallela, Rahul Kumar
March 2003 **The Journal of Machine Learning Research**, Volume 3
Publisher: MIT Press
Full text available: pdf(171.07 KB) Additional Information: full citation, abstract, citings, index terms

High dimensionality of text can be a deterrent in applying complex learners such as Support Vector Machines to the task of text classification. Feature clustering is a powerful alternative to feature selection for reducing the dimensionality of text data. In this paper we propose a new information-theoretic divisive algorithm for feature/word clustering and apply it to text classification. Existing techniques for such "distributional clustering" of words are agglomerative in nature and result in ...

32 Special issue on special feature: An introduction to variable and feature selection
Isabelle Guyon, André Elisseeff
March 2003 **The Journal of Machine Learning Research**, Volume 3
Publisher: MIT Press
Full text available: pdf(862.82 KB) Additional Information: full citation, abstract, citings, index terms

Variable and feature selection have become the focus of much research in areas of application for which datasets with tens or hundreds of thousands of variables are available. These areas include text processing of internet documents, gene expression array analysis, and combinatorial chemistry. The objective of variable selection is three-fold: improving the prediction performance of the predictors, providing faster and more cost-effective predictors, and providing a better understanding of the ...

33 Distributional clustering of English words
Fernando Pereira, Naftali Tishby, Lillian Lee
June 1993 **Proceedings of the 31st annual meeting on Association for Computational Linguistics**
Publisher: Association for Computational Linguistics
Full text available: pdf(756.61 KB) Publisher Site Additional Information: full citation, abstract, references, citings

We describe and evaluate experimentally a method for clustering words according to their distribution in particular syntactic contexts. Words are represented by the relative frequency distributions of contexts in which they appear, and relative entropy between those distributions is used as the similarity measure for clustering. Clusters are represented by average context distributions derived from the given words according to their probabilities of cluster membership. In many cases, the cluster ...

34 Improvements in automatic thesaurus extraction
James R. Curran, Marc Moens
July 2002 **Proceedings of the ACL-02 workshop on Unsupervised lexical acquisition - Volume 9**
Publisher: Association for Computational Linguistics
Full text available: pdf(205.16 KB) Additional Information: full citation, abstract, references

The use of semantic resources is common in modern NLP systems, but methods to extract lexical semantics have only recently begun to perform well enough for practical use. We evaluate existing and new similarity metrics for thesaurus extraction, and experiment with the trade-off between extraction performance and efficiency. We propose an approximation algorithm, based on *canonical attributes* and coarse- and fine-grained matching, that reduces the time complexity and execution time of thes ...

35 Data mining: A matrix density based algorithm to hierarchically co-cluster documents and words
Bhushan Mandhani, Sachindra Joshi, Krishna Kummamuru
May 2003 **Proceedings of the 12th international conference on World Wide Web**
Publisher: ACM Press
Full text available: pdf(133.06 KB) Additional Information: full citation, abstract, references, citings, index terms

This paper proposes an algorithm to hierarchically cluster documents. Each cluster is actually a cluster of documents and an associated cluster of words, thus a document-word co-cluster. Note that, the vector model for documents creates the document-word matrix, of which every co-cluster is a submatrix. One would intuitively expect a submatrix made up of high values to be a good document cluster, with the corresponding word cluster containing its most distinctive features. Our algorithm looks to ...

36 Research track posters: A generalized maximum entropy approach to bregman co-

clustering and matrix approximation
Arindam Banerjee, Inderjit Dhillon, Joydeep Ghosh, Srujana Merugu, Dharmendra S. Modha
August 2004 **Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining KDD '04**
Publisher: ACM Press
Full text available: pdf(166.70 KB) Additional Information: full citation, abstract, references, index terms

Co-clustering is a powerful data mining technique with varied applications such as text clustering, microarray analysis and recommender systems. Recently, an information-theoretic co-clustering approach applicable to empirical joint probability distributions was proposed. In many situations, co-clustering of more general matrices is desired. In this paper, we present a substantially generalized co-clustering framework wherein any Bregman divergence can be used in the objective function, and vari ...

Keywords: Bregman divergences, co-clustering, matrix approximation

37 Multidocument summarization: An added value to clustering in interactive retrieval
Manuel J. Maña-López, Manuel De Buenaga, José M. Gómez-Hidalgo
April 2004 **ACM Transactions on Information Systems (TOIS)**, Volume 22 Issue 2
Publisher: ACM Press
Full text available: pdf(199.91 KB) Additional Information: full citation, abstract, references, index terms, review

A more and more generalized problem in effective information access is the presence in the same corpus of multiple documents that contain similar information. Generally, users may be interested in locating, for a topic addressed by a group of similar documents, one or several particular aspects. This kind of task, called instance or aspectual retrieval, has been explored in several TREC Interactive Tracks. In this article, we propose in addition to the classification capacity of clustering techn ...

Keywords: Multidocument summarization, topic segmentation

38 Research track posters: An objective evaluation criterion for clustering
Arindam Banerjee, John Langford
August 2004 **Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining KDD '04**
Publisher: ACM Press
Full text available: pdf(162.87 KB) Additional Information: full citation, abstract, references, index terms

We propose and test an objective criterion for evaluation of clustering performance: How well does a clustering algorithm run on unlabeled data aid a classification algorithm? The accuracy is quantified using the PAC-MDL bound [3] in a semisupervised setting. Clustering algorithms which naturally separate the data according to (hidden) labels with a small number of clusters perform well. A simple extension of the argument leads to an objective model selection method. Experimental results on text ...

Keywords: MDL, PAC bounds, clustering, evaluation

39 A needle in a haystack: local one-class optimization
Koby Crammer, Gal Chechik
July 2004 **Proceedings of the twenty-first international conference on Machine learning ICML '04**
Publisher: ACM Press
Full text available: pdf(175.88 KB) Additional Information: full citation, abstract, references

This paper addresses the problem of finding a small and coherent subset of points in a given data. This problem, sometimes referred to as *one-class* or *set covering*, requires to find a small-radius ball that covers as many data points as possible. It rises naturally in a

wide range of applications, from finding gene-modules to extracting documents' topics, where many data points are irrelevant to the task at hand, or in applications where only positive examples are available. Most p ...

40 An unsupervised method for multilingual word sense tagging using parallel corpora: a preliminary investigation

Mona Diab

April 2000 **Proceedings of the ACL-2000 workshop on Word senses and multi-linguality - Volume 8**

Publisher: Association for Computational Linguistics

Full text available: pdf(797.04 KB) Additional Information: full citation, abstract, references

With an increasing number of languages making their way to our desktops everyday via the Internet, researchers have come to realize the lack of linguistic knowledge resources for scarcely represented/studied languages. In an attempt to bootstrap some of the required linguistic resources for some of those languages, this paper presents an unsupervised method for automatic multilingual word sense tagging using parallel corpora. The method is evaluated on the English Brown corpus and its translatio ...

Keywords: alignments, multilingual, parallel corpora, unsupervised, word sense tagging

Terms used **cluster** and **margin** and **unsupervised** Found **3,393** of **175,083**

Sort results by: relevance
Display results: expanded form
Save results to a Binder
Search Tips
Open results in a new window

Try an Advanced Search
Try this search in The ACM Guide

Results 41 - 60 of 200 Result page: previous 1 2 **3** 4 5 6 7 8 9 10 next
Best 200 shown Relevance scale

41 Cluster-based find and replace
Robert C. Miller, Alisa M. Marshall
April 2004 **Proceedings of the SIGCHI conference on Human factors in computing systems**
Publisher: ACM Press
Full text available: pdf(190.25 KB) Additional Information: full citation, abstract, references, index terms

In current text editors, the find & replace command offers only two options: replace one match at a time prompting for confirmation, or replace all matches at once without any confirmation. Both approaches are prone to errors. This paper explores a third way: *cluster-based find & replace*, in which the matches are clustered by similarity and whole clusters can be replaced at once. We hypothesized that cluster-based find & replace would make find & replace tasks both faster and more accurat ...

Keywords: clustering, error prevention, find & replace, text editing

42 A unified framework for model-based clustering
Shi Zhong, Joydeep Ghosh
December 2003 **The Journal of Machine Learning Research**, Volume 4
Publisher: MIT Press
Full text available: pdf(851.48 KB) Additional Information: full citation, abstract, index terms

Model-based clustering techniques have been widely used and have shown promising results in many applications involving complex data. This paper presents a unified framework for probabilistic model-based clustering based on a bipartite graph view of data and models that highlights the commonalities and differences among existing model-based clustering algorithms. In this view, clusters are represented as probabilistic models in a model space that is conceptually separate from the data space. For ...

43 Trajectory clustering with mixtures of regression models
Scott Gaffney, Padhraic Smyth
August 1999 **Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining**
Publisher: ACM Press
Full text available: pdf(1.31 MB) Additional Information: full citation, references, citings, index terms

44 MEGA---the maximizing expected generalization algorithm for learning complex query concepts
Edward Chang, Beitao Li

October 2003 **ACM Transactions on Information Systems (TOIS)**, Volume 21 Issue 4
Publisher: ACM Press
Full text available: pdf(1.34 MB) Additional Information: full citation, abstract, references, citings, index terms

Specifying exact query concepts has become increasingly challenging to end-users. This is because many query concepts (e.g., those for looking up a multimedia object) can be hard to articulate, and articulation can be subjective. In this study, we propose a query-concept learner that learns query criteria through an intelligent sampling process. Our concept learner aims to fulfill two primary design objectives: (1) it has to be expressive in order to model most practical query concepts and (2) i ...

Keywords: Active learning, data mining, query concept, relevance feedback

45 Special issue on special feature: Sufficient dimensionality reduction
Amir Globerson, Naftali Tishby
March 2003 **The Journal of Machine Learning Research**, Volume 3
Publisher: MIT Press
Full text available: pdf(266.18 KB) Additional Information: full citation, abstract, citings, index terms

Dimensionality reduction of empirical co-occurrence data is a fundamental problem in unsupervised learning. It is also a well studied problem in statistics known as the analysis of cross-classified data. One principled approach to this problem is to represent the data in low dimension with minimal loss of (mutual) information contained in the original data. In this paper we introduce an information theoretic nonlinear method for finding such a most informative dimension reduction. In contrast wi ...

46 Specialized parsing and grammar induction: A generative constituent-context model for improved grammar induction
Dan Klein, Christopher D. Manning
July 2001 **Proceedings of the 40th Annual Meeting on Association for Computational Linguistics ACL '02**
Publisher: Association for Computational Linguistics
Full text available: pdf(175.36 KB) Additional Information: full citation, abstract, references

We present a generative distributional model for the unsupervised induction of natural language syntax which explicitly models constituent yields and contexts. Parameter search with EM produces higher quality analyses than previously exhibited by unsupervised systems, giving the best published un-supervised parsing results on the ATIS corpus. Experiments on Penn treebank sentences of comparable length show an even higher F1 of 71% on non-trivial brackets. We compare distributionally i ...

47 KM-1 (knowledge management): clustering I: Using bi-modal alignment and clustering techniques for documents and speech thematic segmentations
Dalila Mekhaldi, Denis Lalanne, Rolf Ingold
November 2004 **Proceedings of the thirteenth ACM international conference on Information and knowledge management CIKM '04**
Publisher: ACM Press
Full text available: pdf(463.76 KB) Additional Information: full citation, abstract, references, index terms

In this paper, we describe a new method for a simultaneous thematic segmentation of the meeting dialogs and the documents discussed or visible throughout the meeting. This bi-modal method is suitable for multimodal applications that are centered on documents, such as meetings and lectures, where documents can be aligned with meeting dialogs. Bringing into play this alignment, our bi-modal segmentation method first transforms its results into a set of nodes in a 2D graph space, where the two a ...

Keywords: k-means clustering, thematic alignment, thematic segmentation

48 Clustering on the Unit Hypersphere using von Mises-Fisher Distributions

Arindam Banerjee, Inderjit S. Dhillon, Joydeep Ghosh, Suvrit Sra

September 2005 **The Journal of Machine Learning Research**, Volume 6

Publisher: MIT Press

Full text available: pdf(295.34 KB) Additional Information: full citation, abstract

Several large scale data mining applications, such as text categorization and gene expression analysis, involve high-dimensional data that is also inherently directional in nature. Often such data is L_2 normalized so that it lies on the surface of a unit hypersphere. Popular models such as (mixtures of) multi-variate Gaussians are inadequate for characterizing such data. This paper proposes a generative mixture-model approach to clustering directional data based on the von Mise ...

49 Industry/government track poster: A multinomial clustering model for fast simulation of computer architecture designs

Kaushal Sanghai, Ting Su, Jennifer Dy, David Kaeli

August 2005 **Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining KDD '05**

Publisher: ACM Press

Full text available: pdf(602.87 KB) Additional Information: full citation, abstract, references, index terms

Computer architects utilize simulation tools to evaluate the merits of a new design feature. The time needed to adequately evaluate the tradeoffs associated with adding any new feature has become a critical issue. Recent work has found that by identifying *execution phases* present in common workloads used in simulation studies, we can apply clustering algorithms to significantly reduce the amount of time needed to complete the simulation. Our goal in this paper is to demonstrate the value ...

Keywords: EM, clustering, k-means, mixture of multinomials, program phase, simulation

50 Text classification: Enhanced word clustering for hierarchical text classification

Inderjit S. Dhillon, Subramanyam Mallela, Rahul Kumar

July 2002 **Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining**

Publisher: ACM Press

Full text available: pdf(993.07 KB) Additional Information: full citation, abstract, references, citings, index terms

In this paper we propose a new information-theoretic divisive algorithm for word clustering applied to text classification. In previous work, such "distributional clustering" of features has been found to achieve improvements over feature selection in terms of classification accuracy, especially at lower number of features [2, 28]. However the existing clustering techniques are agglomerative in nature and result in (i) sub-optimal word clusters and (ii) high computational cost. In order to expli ...

51 WaveCluster: a wavelet-based clustering approach for spatial data in very large databases

Gholamhosein Sheikholeslami, Surojit Chatterjee, Aidong Zhang

February 2000 **The VLDB Journal — The International Journal on Very Large Data Bases**, Volume 8 Issue 3-4

Publisher: Springer-Verlag New York, Inc.

Full text available: pdf(594.51 KB) Additional Information: full citation, abstract, citings, index terms

Many applications require the management of spatial data in a multidimensional feature space. Clustering large spatial databases is an important problem, which tries to find the densely populated regions in the feature space to be used in data mining, knowledge discovery, or efficient information retrieval. A good clustering approach should be efficient and detect clusters of arbitrary shape. It must be insensitive to the noise (outliers) and the order of input data. We propose *WaveCluster*

52 Support vector machines: hype or hallelujah?

Kristin P. Bennett, Colin Campbell
December 2000 **ACM SIGKDD Explorations Newsletter**, Volume 2 Issue 2
Publisher: ACM Press
Full text available: pdf(1.26 MB) Additional Information: full citation, citings, index terms

Keywords: Support Vector Machines, kernel methods, statistical learning theory

53 Research track poster: Optimizing time series discretization for knowledge discovery

Fabian Mörchen, Alfred Ultsch
August 2005 **Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining KDD '05**
Publisher: ACM Press
Full text available: pdf(1.34 MB) Additional Information: full citation, abstract, references, index terms

Knowledge Discovery in time series usually requires symbolic time series. Many discretization methods that convert numeric time series to symbolic time series ignore the temporal order of values. This often leads to symbols that do not correspond to states of the process generating the time series and cannot be interpreted meaningfully. We propose a new method for meaningful unsupervised discretization of numeric time series called Persist. The algorithm is based on the Kullback-Leibler divergen ...

Keywords: discretization, persistence, time series

54 Survey articles: Data mining for hypertext: a tutorial survey

Soumen Chakrabarti
January 2000 **ACM SIGKDD Explorations Newsletter**, Volume 1 Issue 2
Publisher: ACM Press
Full text available: pdf(1.19 MB) Additional Information: full citation, abstract, references, citings

With over 800 million pages covering most areas of human endeavor, the World-wide Web is a fertile ground for data mining research to make a difference to the effectiveness of information search. Today, Web surfers access the Web through two dominant interfaces: clicking on hyperlinks and searching via keyword queries. This process is often tentative and unsatisfactory. Better support is needed for expressing one's information need and dealing with a search result in more structured ways than av ...

55 Industry/government track posters: Learning a complex metabolomic dataset using random forests and support vector machines

Young Truong, Xiaodong Lin, Chris Beecher
August 2004 **Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining KDD '04**
Publisher: ACM Press
Full text available: pdf(179.85 KB) Additional Information: full citation, abstract, references, index terms

Metabolomics is the "omics" science of biochemistry. The associated data include the quantitative measurements of all small molecule metabolites in a biological sample. These datasets provide a window into dynamic biochemical networks and conjointly with other "omic" data, genes and proteins, have great potential to unravel complex human diseases. The dataset used in this study has 63 individuals, normal and diseased, and the diseased are drug treated or not, so there are three classes. The goal ...

Keywords: metabolomics, missing data, random forest, support vector machines

56 Web page classification: PEBL: positive example based learning for Web page

classification using SVM
Hwanjo Yu, Jiawei Han, Kevin Chen-Chuan Chang
July 2002 **Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining**
Publisher: ACM Press
Full text available: pdf(1.01 MB) Additional Information: full citation, abstract, references, citings, index terms

Web page classification is one of the essential techniques for Web mining. Specifically, classifying Web pages of a user-interesting class is the first step of mining interesting information from the Web. However, constructing a classifier for an interesting class requires laborious pre-processing such as collecting positive and negative training examples. For instance, in order to construct a "homepage" classifier, one needs to collect a sample of homepages (positive examples) and a sample of n ...

Keywords: Mapping-Convergence (M-C) algorithm, SVM (Support Vector Machine), labeled data, unlabeled data

57 Extracting predicates from mining models for efficient query evaluation
Surajit Chaudhuri, Vivek Narasayya, Sunita Sarawagi
September 2004 **ACM Transactions on Database Systems (TODS)**, Volume 29 Issue 3
Publisher: ACM Press
Full text available: pdf(698.37 KB) Additional Information: full citation, abstract, references, index terms

Modern relational database systems are beginning to support ad hoc queries on mining models. In this article, we explore novel techniques for optimizing queries that contain predicates on the results of application of mining models to relational data. For such queries, we use the internal structure of the mining model to automatically derive traditional database predicates. We present algorithms for deriving such predicates for a large class of popular discrete mining models: decision trees, nai ...

Keywords: Complex predicate optimization, simpler rules from complex predictive functions

58 Content 6: multimodal processing: Graph based multi-modality learning
Hanghang Tong, Jingrui He, Mingjing Li, Changshui Zhang, Wei-Ying Ma
November 2005 **Proceedings of the 13th annual ACM international conference on Multimedia MULTIMEDIA '05**
Publisher: ACM Press
Full text available: pdf(303.64 KB) Additional Information: full citation, abstract, references, index terms

To better understand the content of multimedia, a lot of research efforts have been made on how to learn from multi-modal feature. In this paper, it is studied from a graph point of view: each kind of feature from one modality is represented as one independent graph; and the learning task is formulated as inferring from the constraints in every graph as well as supervision information (if available). For semi-supervised learning, two different fusion schemes, namely linear form and sequential fo ...

Keywords: Bayesian interpretation, graph model, multi-modality analysis, regularized optimization, similarity propagation

59 Regular papers: Using a probabilistic class-based lexicon for lexical ambiguity resolution
Detlef Prescher, Stefan Riezler, Mats Rooth
July 2000 **Proceedings of the 18th conference on Computational linguistics - Volume 2**
Publisher: Association for Computational Linguistics
Full text available: Additional Information:

pdf(649.30 KB) full citation, abstract, references

This paper presents the use of probabilistic class-based lexica for disambiguation in target-word selection. Our method employs minimal but precise contextual information for disambiguation. That is, only information provided by the target-verb, enriched by the condensed information of a probabilistic class-based lexicon, is used. Induction of classes and fine-tuning to verbal arguments is done in an unsupervised manner by EM-based clustering techniques. The method shows promising results in an ...

60 Core Vector Machines: Fast SVM Training on Very Large Data Sets

Ivor W. Tsang, James T. Kwok, Pak-Ming Cheung

September 2005 **The Journal of Machine Learning Research**, Volume 6

Publisher: MIT Press

Full text available: pdf(417.31 KB) Additional Information: full citation, abstract

Standard SVM training has $O(m^3)$ time and $O(m^2)$ space complexities, where m is the training set size. It is thus computationally infeasible on very large data sets. By observing that practical SVM implementations only *approximate* the optimal solution by an iterative strategy, we scale up kernel methods by exploiting such "approximateness" in this paper. We first show that many kernel methods can be equivalently formulated as minimum ...

Web Results **1** - **10** of about **44,600** for **+cluster +margin +unsupervised**. (**0.09** seconds)

[Paper] **Unsupervised** MRI Tissue Classification by Support Vector ...
Neither misclassification nor sam- ples inside the **margin** are allowed. ... If a **cluster** had pixels representing more than one tissue, all the pixels in the ...
www.actapress.com/PDFViewer.aspx?paperId=16305 - Similar pages

[PDF] **Unsupervised** and Semi-supervised Multi-class Support Vector Machines
File Format: PDF/Adobe Acrobat - View as HTML
ming can be used for **unsupervised** training of two-class. SVMs (Xu et al. ... addition, the slack parameter for maximum **margin cluster**- ...
www.cs.ualberta.ca/~dale/papers/aaai05.pdf - Similar pages

[PDF] Discriminative Clustering of Yeast Stress Response
File Format: PDF/Adobe Acrobat - View as HTML
is the **cluster margin**, that is, number of data samples in ... **unsupervised** clustering is not enough; it is particularly interesting to search ...
www.cis.hut.fi/jnikkila/papers/bcip05.pdf - Similar pages

[PDF] **Unsupervised** Document Classification using Sequential Information ...
File Format: PDF/Adobe Acrobat - View as HTML
ically by a significant **margin**. Moreover, the sIB results are ... ing **cluster**'s recall to gain higher precision, and show how ...
www.cs.rutgers.edu/~mlittman/topics/dimred02/slonim.pdf - Similar pages

[PDF] **Cluster** analysis of BI-RADS descriptions of biopsy-proven breast ...
File Format: PDF/Adobe Acrobat - View as HTML
Recall that this analysis was performed in an **unsupervised**. fashion and that the clustering ... Mass **Margin**. **Cluster**. Mass Shape. Proc. SPIE Vol. 4684 ...
www.bme.utexas.edu/research/ informatics/pubs/2002SPIE.pdf - Similar pages

Project-Team-Imedia:Clustering and learning
While it is easy to apply standard **unsupervised** clustering algorithms to the ... Following this remarks, we consider that the least well-defined **cluster** at ...
www.inria.fr/rapportsactivite/RA2005/imedia/uid91.html - 42k - Cached - Similar pages

Unsupervised document classification using sequential information ...
We apply this algorithm to **unsupervised** document classification. ... Finally, we propose a simple procedure for trading **cluster's** recall to gain higher ...
portal.acm.org/ citation.cfm?coll=GUIDE&dl=GUIDE&id=564401 - Similar pages

[PDF] **Cluster**-Based Find and Replace
File Format: PDF/Adobe Acrobat - View as HTML
Clustering, also called **unsupervised** learning, has a long ... when the mouse was over the **cluster's margin**. This re-. duced but didn't eliminate the problem ...
people.csail.mit.edu/rcm/chi04.pdf - Similar pages

[PDF] **Unsupervised** Clustering of Images using their Joint Segmentation ...
File Format: PDF/Adobe Acrobat - View as HTML
words/documents, to apply algorithms from the field of **unsupervised** document. classification to **cluster** the images. The first step may be regarded as ...
www.cs.huji.ac.il/~seldin/publications/SSW_SCTV03.pdf - Similar pages

[PDF] Name Discrimination and Email Clustering using **Unsupervised** ...
File Format: PDF/Adobe Acrobat

adapting the **unsupervised** word sense discrimination methods. The main objec-. tive is to be able to **cluster** a given set of emails based upon the overall ...
www.d.umn.edu/~kulka020/iicai05-kulkarni.pdf - Similar pages

Gooooooooogle ▶
Result Page: 1 2 3 4 5 6 7 8 9 10 **Next**

+cluster +margin +unsupervised Search

Search within results | Language Tools | Search Tips | Dissatisfied? Help us improve

Google Home - Advertising Programs - Business Solutions - About Google

Web Results **11** - **20** of about **44,600** for **+cluster +margin +unsupervised**. (**0.26** seconds)

[PPT] Support Vector Clustering
File Format: Microsoft Powerpoint 97 - View as HTML
Unsupervised Learning - There is no training process involved. Eg Clustering. ... Where nbsv is the number of BSVs and C is soft **margin** constant ...
www.cse.psu.edu/~datta/Present/svc.ppt - Similar pages

[PPT] Data Mining A Tutorial-Based Primer
File Format: Microsoft Powerpoint - View as HTML
Unsupervised clustering. Assume 3 clusters; Representative rules from each **cluster**; IF **Margin** Account=yes & Age=20-29 & Annual Income=40-59K. THEN **Cluster**=1 ...
www.doc.gold.ac.uk/~mas01ds/ cis338/lectures/lecture1.ppt - Similar pages

Cancer characterization and feature set extraction by ...
The authors use a completely **unsupervised** self organizing map technique to **cluster** ... **Margin** for the test samples along with the label of the **cluster** which ...
www.pubmedcentral.nih.gov/ articlerender.fcgi?artid=385290 - Similar pages

[PDF] **Unsupervised** Document Classification using Sequential Information ...
File Format: PDF/Adobe Acrobat - View as HTML
ically by a significant **margin**. Moreover, the sIB results are ... any given agglomerative procedure into a sequential **cluster**-. ing algorithm. ...
www.princeton.edu/~nslonim/publications/SIGIR2002.pdf - Similar pages

[PS] Semi-Supervised Clustering Using Genetic Algorithms Ayhan Demiriz ...
File Format: Adobe PostScript - View as HTML
The objective function of an **unsupervised** technique, eg K-means clustering, is modified to minimize both the within **cluster** variance of the input attributes ...
www.rpi.edu/~bennek/annie.ps - Similar pages

[PDF] Revealing Predictive Gene Clusters with Supervised Algorithms
File Format: PDF/Adobe Acrobat - View as HTML
All these methods **cluster** genes according to **unsupervised** similarity mea- ... The criterion is refined with a second priority **margin** function M, ...
www.ci.tuwien.ac.at/Conferences/ DSC-2003/Proceedings/Dettling.pdf - Similar pages

> [PDF] Revealing Predictive Gene Clusters with Supervised Algorithms
> File Format: PDF/Adobe Acrobat - View as HTML
> least) into the current **cluster** in terms of an **unsupervised** similarity ... with a second priority **margin** function M, measuring the size of the gap (in stan- ...
> www.ci.tuwien.ac.at/Conferences/ DSC-2003/Drafts/Dettling.pdf - Similar pages

[PDF] Learning intrusion detection: supervised or **unsupervised**?
File Format: PDF/Adobe Acrobat - View as HTML
The process is repeated until the **cluster** centers do not ... best results (with a significant **margin**) are attained by the SVM, which can be ...
ida.first.fraunhofer.de/~rieck/docs/iciap2005.pdf - Similar pages

[PDF] **Unsupervised** improvement of visual detectors using co-training ...
File Format: PDF/Adobe Acrobat
EM is typically used to infer missing **cluster** labels. During ... gin" and attempt to maximize the **margin** of all (or most). training examples. ...
ieeexplore.ieee.org/iel5/8769/27772/ 01238406.pdf?tp=&arnumber=1238406&isnumber=27772 - Similar pages

[PPT] Slide 1
File Format: Microsoft Powerpoint 97 - View as HTML
Unsupervised Clustering: Partitioning Methods. K-means Algorithm partitions a set of n objects into k clusters so that the resulting intra-**cluster** ...
genome.osu.edu/ibgp730/lectures_2003/Clustering.ppt - Similar pages

◀ Goooooooooogle ▶

Result Page: **Previous** 1 **2** 3 4 5 6 7 8 9 1011 **Next**

+cluster +margin +unsupervised Search

Search within results | Language Tools | Search Tips

Google Home - Advertising Programs - Business Solutions - About Google

Web Results **21 - 30** of about **44,600** for **+cluster +margin +unsupervised**. (**0.18** seconds)

[PDF] Supervised Clustering of Genes
File Format: PDF/Adobe Acrobat - View as HTML
we neither know **cluster** size, the number of clusters q, nor the function f(·) ... S is the value of the Wilcoxon test statistic, refined by the **margin** ...
www.bgx.org.uk/wye/talks/dettling.pdf - Similar pages

[PS] Using Competitive Learning to Handle Missing Values in ...
File Format: Adobe PostScript - View as HTML
In this paper, we address the **unsupervised** clustering of galaxies, concentrating on ...
Cluster 0 1 2 3 4 5 6 7 8 9 10 Elliptical 9 10 5 11 4 47 56 1 7 0 2 ...
www.cs.queensu.ca/TechReports/Reports/2002-458.ps - Similar pages

[PDF] **Unsupervised** Evidence Integration
File Format: PDF/Adobe Acrobat - View as HTML
also be thought of as the problem faced by **cluster**- ... tical method of **unsupervised** evidence integration in- ... This maximum **margin** might be viewed as ...
www.machinelearning.org/proceedings/
icml2005/papers/066_EvidenceIntegration_LongEtAl.pdf - Similar pages

Genome Biology | Full text | Supervised clustering of genes
For every permuted set of responses, a single **cluster** (q = 1) was formed on the entire dataset and both its final score s*(l) and **margin** m*(l) were recorded ...
genomebiology.com/2002/3/12/research/0069 - 147k - Cached - Similar pages

[PDF] **Cluster** Validity Through Graph-based Boundary Analysis
File Format: PDF/Adobe Acrobat - View as HTML
der points from each **cluster**. •. Compare the **margin** to the local neighborhood dis- ...
Parametric and non-parametric **unsupervised**. **cluster** analysis. ...
www.cs.jcu.edu.au/ftp/pub/techreports/Validity.pdf - Similar pages

BioMed Central | Full text | Cancer characterization and feature ...
We run the discriminative **margin** clustering procedure on the set of tumor samples ...
unsupervised self organizing map technique to **cluster** gene expression ...
www.biomedcentral.com/1471-2105/5/21 - 109k - Cached - Similar pages

[PDF] AM Bagirov, AM Rubinov, NV Soukhoroukova and J. Yearwood ...
File Format: PDF/Adobe Acrobat
the maximal **margin** of separation between the two classes comes from the ...
Unsupervised and Supervised Data Classification. 35. c. n. **cluster** ...
top.umh.es/top11101.pdf - Similar pages

An Improved **Cluster** Labeling Method for Support Vector Clustering
... recently emerged **unsupervised** learning method inspired by support vector machines. ...
A new **cluster** labeling method for SVC is developed based on some ...
doi.ieeecomputersociety.org/10.1109/TPAMI.2005.47 - Similar pages

[PDF] Automated Design and Discovery of Novel
File Format: PDF/Adobe Acrobat - View as HTML
The objective function of an **unsupervised** technique, eg K-. means clustering, is modified to minimize both the within **cluster** variance of the input ...
https:/.../other/samba/2.0.5/common/public/
locker/82/001182/public_html/files/pdf_files/progress_report.pdf - Similar pages

[PPT] Slide 1
File Format: Microsoft Powerpoint 97 - View as HTML
Similarly to SVMs, train i so that **margin** is largest for 0 i 1 ... Attempt to "contract" the distances within each **cluster** while keeping ...
ai.stanford.edu/~serafim/CS374_2005/ Presentations/Lecture14_ProteinClassification.ppt - Similar pages

◀ Goooooooooooogle ▶
Result Page: **Previous** 1 2 3 4 5 6 7 8 9 10 11 12 **Next**

+cluster +margin +unsupervised Search

Search within results | Language Tools | Search Tips

Google Home - Advertising Programs - Business Solutions - About Google

USPTO Patent Full-Text and Image Database

Searching US Patent Collection...

Results of Search in US Patent Collection db for:
((cluster AND margin) AND unsupervised): 27 patents.
Hits 1 through 27 out of 27

Jump To

Refine Search cluster and margin and unsupervised

	PAT. NO.	Title
1	7,031,936	Methods and systems for automated inferred valuation of credit scoring
2	7,028,005	Methods and systems for finding value and reducing risk
3	7,016,882	Method and apparatus for evolutionary design
4	7,003,484	Methods and systems for efficiently sampling portfolios for optimal underwriting
5	6,985,881	Methods and apparatus for automated underwriting of segmentable portfolio assets
6	6,949,342	Prostate cancer diagnosis and outcome prediction by expression analysis
7	6,941,287	Distributed hierarchical evolutionary modeling and visualization of empirical data
8	6,904,408	Bionet method, system and personalized web content manager responsive to browser viewers' psychological preferences, behavioral responses and physiological stress indicators
9	6,882,997	Wavelet-based clustering method for managing spatial data in very large databases
10	6,882,990	Methods of identifying biological patterns using multiple data sets
11	6,862,710	Internet navigation using soft hyperlinks
12	6,789,069	Method for enhancing knowledge discovered from biological data using a learning machine
13	6,785,672	Methods and apparatus for performing sequence homology detection
14	6,760,715	Enhancing biological knowledge discovery using multiples support vector machines
15	6,757,646	Extended functionality for an inverse inference engine based web search
16	6,754,589	System and method for enhanced hydrocarbon recovery
17	6,714,925	System for identifying patterns in biological data using a distributed network
18	6,574,565	System and method for enhanced hydrocarbon recovery
19	6,571,199	Method and apparatus for performing pattern dictionary formation for use in sequence homology detection
20	6,510,406	Inverse inference engine for high performance web search
21	6,411,903	System and method for delineating spatially dependent objects, such as hydrocarbon accumulations from seismic data

22 6,236,942 [T] System and method for delineating spatially dependent objects, such as hydrocarbon accumulations from seismic data

23 6,119,112 [T] Optimum cessation of training in neural networks

24 5,625,751 [T] Neural network for contingency ranking dynamic security indices for use under fault conditions in a power distribution system

25 5,590,218 [T] Unsupervised neural network classification with back propagation

26 5,485,908 [T] Pattern recognition using artificial neural network for coin validation

27 5,444,796 [T] Method for unsupervised neural network classification with back propagation

---